ABSTRAK

Deffan Rizki Nanda Prayooga. 2026. Optimasi *Hyperparameter Random Forest* dengan *Stratified K-Fold Cross Validation* untuk Klasifikasi Tingkat Stres Mahasiswa. Skripsi. Program Sarjana. Program Studi Teknik Informatika. Universitas Muhammadiyah Jember.

Pembimbing: Moh. Dasuki, M.Kom; Triawan Adi Cahyanto, M.Kom.

Tingkat stres mahasiswa merupakan salah satu aspek yang perlu diperhatikan karena dapat dipengaruhi oleh berbagai faktor psikologis, akademik, sosial, dan lingkungan. Penelitian ini bertujuan untuk mengklasifikasikan tingkat stres mahasiswa menggunakan algoritma *Random Forest* serta membandingkan performa model sebelum dan sesudah optimasi *hyperparameter*. Data yang digunakan adalah *dataset Student Stress Factors: A Comprehensive Analysis* yang terdiri atas 1.100 data, 20 fitur, dan tiga kategori tingkat stres. Optimasi *hyperparameter* dilakukan menggunakan *GridSearchCV* dengan penerapan *Nested Stratified K-Fold Cross Validation*. Evaluasi model menggunakan metrik Accuracy, Precision Macro, Recall Macro, dan F1-Score Macro, sedangkan uji McNemar digunakan untuk menganalisis perbedaan kesalahan klasifikasi kedua model. Hasil penelitian menunjukkan bahwa model *baseline* memperoleh Accuracy sebesar 87,6727%, sedangkan model *optimized* memperoleh 87,6909%. Hasil uji McNemar menunjukkan nilai p lebih besar dari 0,05 pada seluruh pengujian berdasarkan random seed, sehingga belum terdapat bukti statistik yang cukup untuk menyatakan adanya perbedaan kesalahan klasifikasi yang signifikan antara kedua model. Dengan demikian, optimasi *hyperparameter* menghasilkan peningkatan nilai Accuracy yang sangat kecil dan belum menunjukkan perbedaan performa yang signifikan secara statistik.

Kata kunci: *Random Forest*, optimasi *hyperparameter*, klasifikasi tingkat stres mahasiswa, *GridSearchCV*, *Nested Cross Validation*.

ABSTRACT

Deffan Rizki Nanda Prayoga. *2026. Random Forest Hyperparameter Optimization Using Stratified K-Fold Cross-Validation for Student Stress Level Classification. Undergraduate Thesis. Undergraduate Program. Informatics Engineering Study Program. University of Muhammadiyah Jember.*

***Advisors** : Moh. Dasuki, M.Kom; Triawan Adi Cahyanto, M.Kom.*

Student stress is an important issue to consider because it can be influenced by various psychological, academic, social, and environmental factors. This study aims to classify student stress levels using the Random Forest algorithm and compare the performance of baseline and optimized models. The dataset used was Student Stress Factors: A Comprehensive Analysis, consisting of 1,100 records, 20 features, and three stress-level categories. Hyperparameter optimization was performed using GridSearchCV within a Nested Stratified K-Fold Cross-Validation framework. Model performance was evaluated using Accuracy, Macro Precision, Macro Recall, and Macro F1-Score, while the McNemar test was used to analyze differences in classification errors between the two models. The results showed that the baseline model achieved an accuracy of 87.6727%, while the optimized model achieved 87.6909%. The McNemar test yielded p-values greater than 0.05 across all random seed evaluations, indicating insufficient statistical evidence of a significant difference in classification errors between the two models. Therefore, hyperparameter optimization resulted in only a marginal improvement in accuracy and did not demonstrate a statistically significant difference in model performance.

***Keywords**: Random Forest, hyperparameter optimization, student stress level classification, GridSearchCV, Nested Cross-Validation.*